Sample variance in weak lensing: How many simulations are required?

Andrea Petri, Zoltán Haiman, and Morgan May

Phys. Rev. D **93**, 063524 — Published 24 March 2016

DOI: 10.1103/PhysRevD.93.063524

Sample variance in weak lensing: how many simulations are required?

Andrea Petri,[1,2,*] Zoltán Haiman,[3] and Morgan May[2]

[1] *Department of Physics, Columbia University, New York, NY 10027, USA*
[2] *Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*
[3] *Department of Astronomy, Columbia University, New York, NY 10027, USA*

Constraining cosmology using weak gravitational lensing consists of comparing a measured feature vector of dimension N_b with its simulated counterpart. An accurate estimate of the $N_b \times N_b$ feature covariance matrix $\mathbf{C}$ is essential to obtain accurate parameter confidence intervals. When $\mathbf{C}$ is measured from a set of simulations, an important question is how large this set should be. To answer this question, we construct different ensembles of N_r realizations of the shear field, using a common randomization procedure that recycles the outputs from a smaller number $N_s \leq N_r$ of independent ray-tracing N–body simulations. We study parameter confidence intervals as a function of (N_s, N_r) in the range $1 \leq N_s \leq 200$ and $1 \leq N_r \lesssim 10^5$. Previous work [1] has shown that Gaussian noise in the feature vectors (from which the covariance is estimated) lead, at quadratic order, to an $O(1/N_r)$ degradation of the parameter confidence intervals. Using a variety of lensing features measured in our simulations, including shear-shear power spectra and peak counts, we show that cubic and quartic covariance fluctuations lead to additional $O(1/N_r^2)$ error degradation that is not negligible when N_r is only a factor of few larger than N_b. We study the large N_r limit, and find that a single, 240Mpc/h sized 512^3-particle N–body simulation ($N_s = 1$) can be repeatedly recycled to produce as many as N_r = few $\times 10^4$ shear maps whose power spectra and high-significance peak counts can be treated as statistically independent. As a result, a small number of simulations ($N_s = 1$ or 2) is sufficient to forecast parameter confidence intervals at percent accuracy.

PACS numbers: 98.80.-k, 95.36.+x, 95.30.Sf, 98.62.Sb
Keywords: Weak Gravitational Lensing — Simulations — Methods: analytical, numerical, statistical

I. INTRODUCTION

Weak gravitational lensing (WL) is a promising cosmological probe for constraining the dark energy equation of state w, and has been considered by a range of past (CFHTLens [2, 3], COSMOS [4]), ongoing (DES [5]) and future (LSST [6], Euclid [7], WFIRST [8]) experiments. In an era where cosmology is data driven, accurate numerical simulations of shear fields are becoming important for several reasons, including assessing baryonic effects [9–14], the utility of non–Gaussian statistics [15–23] and various systematic effects [24–27].

A fundamental issue with predictions from simulations is that the finite number of simulations naturally introduces fluctuations in the forecasts, due to inevitable sample variance [28]. In general, quantities such as the mean or the variance of any feature (e.g. the shear power spectrum at a multipole ℓ), measured from a finite set of simulations, will fluctuate, and can also suffer a bias. While biases in the estimates of both the mean and the variance have been studied extensively, the impact of fluctuations in the variance has received less attention. These fluctuations have been shown to have non-negligible effects on estimates of features covariances and hence on parameter constraints. In particular, in the limit of Gaussian fluctuations, the parameter confidence limits are degraged by a factor $1 + O(1/N_r)$ [1, 29].

This work studies these issues further, focusing on the number of independent N–body simulations required for an accurate estimate of the parameter constraints. Ray-tracing simulations that resolve the cosmic structures responsible for lensing on arcminute scales are limited to physical sizes of hundreds of Mpc, and thus cover a solid angle of only $O(10\ \mathrm{deg}^2)$. As a result, many simulations are required to tile a significant fraction of the sky, and to make predictions for large "all-sky" surveys, such as the ones by DES, LSST, Euclid, WFIRST. In practice, this has led to the wide-spread use of "pseudo-independent" realizations, i.e. a procedure in which one randomizes and re-cycles the output of a single 3D simulation multiple times. In light of the forthcoming large surveys, it is imperative to assess the statistical validity of this approach, and to ask how many times a single simulation can be fairly recycled. In this paper, we address these questions with ensembles of up to $N_r = 10^5$ random realizations, extracted from up to $N_s = 200$ independent ray-tracing N–body simulations. We focus in particular on the parameter w, and on two different statistics: the (convergence) power spectrum and the number counts of peaks.

This paper is organized as follows. In § II, we summarize the shear simulation methods we utilized, and describe the formalism we adopted to forecast cosmological parameter constraints. We then vary the number of simulations and the number of pseudo-independent realizations, and present our main findings in § III. These results are discussed further in § IV. We offer our conclusions, and suggest follow-up future work in § V.

* apetri@phys.columbia.edu

II. METHODS

A. Ray-tracing simulations of the convergence field

In this section, we describe how we constructed our shear field ensembles. Background galaxies at redshift z_s are lensed by large scale structures between $z = 0$ and z_s. The shape distortions due to the cosmic shear $\boldsymbol{\gamma}$ can be computed in terms of the dark matter gravitational potential $\Phi(\mathbf{x}, z)$. Because the evolution of Φ with redshift is non–linear, it needs to be computed with numerical simulations. We make use of the public code `Gadget2` [30], with which we run a sequence of 200 independent dark–matter–only N–body simulations that track the evolution of the density fluctuations. We assume a standard ΛCDM background universe with the parameters $(\Omega_m, \Omega_\Lambda, h, w, \sigma_8, n_s) = (0.26, 0.74, 0.72, -1, 0.8, 0.96)$. We fix the comoving size of the simulation box to 240Mpc/h, and use 512^3 particles, corresponding to a dark matter particle mass of $\approx 10^{10} M_\odot$.

We assume a uniform galaxy distribution at a constant redshift $z_s = 2$ (at which the simulation box has an angular size of $\theta_{\rm box} = 3.5°$) and we discretize the mass distribution between z_s and the observer at $z = 0$ with a sequence of 46 two dimensional lenses of thickness 80Mpc/h. The surface density on each lens plane is computed by projecting the three–dimensional density measured from `Gadget2` snapshots. We then apply the multi–lens–plane algorithm (see [31, 32] for example) to trace the deflections of $n^2_{\rm ray} = 2048^2$ light rays arranged on a square grid of total size $\theta_{\rm box}$, from $z = 0$ to z_s. This corresponds to a pixel angular resolution of $0.1'$. Our implementation of this algorithm is part of the `LensTools` computing package we have been developing [33], and have released under the `MIT` license. Many different realizations r of the same shear field $\boldsymbol{\gamma}_r(\boldsymbol{\theta})$ can be generated by picking different lens planes that lie between the observer and z_s. The randomization procedure we adopt is the following (see [34] for reference):

- For each lens-plane redshift z_l, select the snapshot at z_l from the i–th N–body simulation, where i is a random integer $i \in [1, N_s]$.
- Choose randomly between the three orthogonal directions $\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z$: the lens plane will be perpendicular to this direction.
- Choose the position of the plane along the snapshot: because the lens thickness is 1/3 the size of the box, we can cut three different slices of the simulation box for each orientation $\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z$. This gives a total of 9 choices for generating a lens plane out of a single N–body snapshot.
- Perform a periodic random shift of the lens plane along its two directions.
- Repeat the above procedure for each lens-plane redshift z_l.

This randomization procedure allows us to produce an (almost) arbitrary number N_r of shear realizations $\boldsymbol{\gamma}_r(\boldsymbol{\theta})$. However, these realizations are not guaranteed to be independent, if N_s is not large enough. Using the set of 200 independent N–body simulations, we construct different ensembles with different choices of $N_s \in [1, 200]$. Each of these ensembles consists of the same number $N_r = 1000$ of shear realizations. We also build an additional ensemble with $N_s = 1$ and $N_r = 10^5$ realizations. For each realization of each ensemble, we reconstruct the convergence $\kappa_r(\boldsymbol{\theta})$ from the trace of the light-ray deflection Jacobian matrix, measured from the difference in deflection angles between nearby light-rays [31, 32, 34].

We measure the κ angular power spectrum $P^{\kappa\kappa}_r(\ell)$ defined as

$$\langle \tilde{\kappa}_r(\boldsymbol{\ell}) \tilde{\kappa}_r(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} + \boldsymbol{\ell}') P^{\kappa\kappa}_r(\ell) \tag{1}$$

As an additional summary statistic, we consider the counts of local κ maxima of a certain height κ_0, $n_r(\kappa_0)$ (hereafter *peak counts*), with varying κ_0 chosen between the minimum and maximum values measured from the maps $(\kappa_{\rm min}, \kappa_{\rm max}) = (-0.06, 0.45)$. Different choices of κ_0 binning used in this work are outlined in Table II. The fact that the ensemble of N_r realizations is not completely independent if N_s is not large enough can have an effect on the covariance estimators of both $P^{\kappa\kappa}$ and $n(\kappa_0)$.

To measure the cosmological dependence of the κ peak counts, we performed a set of additional ray–tracing simulations with different combinations of the cosmological parameter triplet (Ω_m, w, σ_8). A summary of the complete set of shear ensembles used in this work is listed in Table I.

B. Cosmological parameter inference

Let $\hat{\mathbf{d}}$ be a single estimate for a feature of dimension N_b, $\mathbf{d}(\mathbf{p})$ be the true value of this feature at a point $\mathbf{p}$ in parameter space (which has a dimension N_p) and $\mathbf{C}$ be the $N_b \times N_b$ feature covariance matrix. For the purpose of this work $\mathbf{p}$ is the triplet (Ω_m, w, σ_8) and $\mathbf{d}$ is one of the features – either a power spectrum or a peak count histogram – in Table II. Although existing emulators can be used, in principle, to compute both $\mathbf{d}(\mathbf{p})$ and $\mathbf{C}$, the latter is more difficult, and typically only the mean, $\mathbf{d}(\mathbf{p})$, has been computed to date (refs. [35, 36], but see an exception by ref. [37]). Estimating $\mathbf{C}$ from simulations involves generating a series of mock realizations $\hat{\mathbf{d}}_r$ with $r = 1...N_r$ and computing the sample covariance $\hat{\mathbf{C}}$,

$$\bar{\mathbf{d}} = \frac{1}{N_r} \sum_{r=1}^{N_r} \hat{\mathbf{d}}_r, \tag{2}$$

$$\hat{\mathbf{C}} = \frac{1}{N_r - 1} \sum_{r=1}^{N_r} (\hat{\mathbf{d}}_r - \bar{\mathbf{d}})(\hat{\mathbf{d}}_r - \bar{\mathbf{d}})^T. \quad (3)$$

Assuming a normal feature likelihood, together with a flat prior on the parameter space, the parameter posterior distribution $\mathcal{L}(\mathbf{p}|\hat{\mathbf{d}}_{\rm obs})$ given an observed instance of $\hat{\mathbf{d}}$, which we call $\hat{\mathbf{d}}_{\rm obs}$, follows from Bayes' theorem,

$$-2 \log \mathcal{L}(\mathbf{p}|\hat{\mathbf{d}}_{\rm obs}) = [\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}(\mathbf{p})]^T \hat{\mathbf{C}}^{-1} [\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}(\mathbf{p})]. \quad (4)$$

For the sake of simplicity, we approximate the posterior as a Gaussian around its maximum. This corresponds to Taylor-expanding the simulated feature to first order around a point $\mathbf{p}_0$ (ideally the maximum of Eq. 4):

$$\mathbf{d}(\mathbf{p}) \approx \mathbf{d}_0 + \mathbf{d}'_0(\mathbf{p} - \mathbf{p}_0). \quad (5)$$

We chose $\mathbf{p}_0$ to be the triplet $(\Omega_m, w, \sigma_8) = (0.26, -1, 0.8)$. To measure the derivatives of the features $\mathbf{d}'_0$ with respect to the cosmological parameters, we make use of the public code `Nicaea` [36] for the power spectrum, and we use an independent simulation set (containing simulations with a variety of different combinations of (Ω_m, w, σ_8), see Table I) for the peak counts.

We can build the estimator for the posterior maximum $\hat{\mathbf{p}}$, given the observation $\hat{\mathbf{d}}_{\rm obs}$, as follows:

$$\hat{\mathbf{p}} = \mathbf{p}_0 + \hat{\mathbf{T}}(\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}_0), \quad (6)$$

$$\hat{\mathbf{T}} = (\mathbf{d}_0'^T \hat{\mathbf{C}}^{-1} \mathbf{d}'_0)^{-1} \mathbf{d}_0'^T \hat{\mathbf{C}}^{-1}. \quad (7)$$

Because $\hat{\mathbf{p}}$ is estimated using a single noisy data instance $\hat{\mathbf{d}}_{\rm obs}$, its estimate will be scattered around the true value $\langle \hat{\mathbf{p}} \rangle_O$. In the following we use the $\langle \rangle_O$ notation for expectation values taken with respect to observations, while we keep the notation $\langle \rangle$ for expectation values taken with respect to the simulations. Defining the *precision matrix* $\hat{\mathbf{\Psi}} = \hat{\mathbf{C}}^{-1}$, we can express the estimator of the observational scatter in $\hat{\mathbf{p}}$:

$$\hat{\Sigma}_{\mathbf{p}} = \hat{\mathbf{F}}^{-1} \mathbf{d}_0'^T \hat{\mathbf{\Psi}} \langle (\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}_0)(\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}_0)^T \rangle_O \hat{\mathbf{\Psi}} \mathbf{d}'_0 \hat{\mathbf{F}}^{-1}, \quad (8)$$

$$\hat{\mathbf{F}} = \mathbf{d}_0'^T \hat{\mathbf{\Psi}} \mathbf{d}'_0. \quad (9)$$

Here we introduced the familiar Fisher matrix estimator $\hat{\mathbf{F}} = \mathbf{d}_0'^T \hat{\mathbf{\Psi}} \mathbf{d}'_0$ and, for simplicity, we assumed $\langle \hat{\mathbf{d}}_{\rm obs} \rangle_O = \mathbf{d}_0$, so that $\langle (\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}_0)(\hat{\mathbf{d}}_{\rm obs} - \mathbf{d}_0)^T \rangle_O = \mathbf{C}$. When we perform an observation $\hat{\mathbf{d}}_{\rm obs}$, the parameter estimate $\hat{\mathbf{p}}$ is a random draw from a probability distribution with variance $\hat{\Sigma}_{\mathbf{p}}$, which inherits noise from the simulations. The noise in the covariance estimator (Eq. 3) and in its inverse $\hat{\mathbf{\Psi}}$ propagate all the way to the posterior (Eq. 4), the parameter estimate (Eq. 6) and its variance (Eq. 8). Following refs. [1, 29, 38] we compute the expectation value of Eq. (8) over simulations, $\langle \hat{\Sigma}_{\mathbf{p}} \rangle$, up to $O(1/N_r^2)$ by expanding Eq. (8) to quartic order in the statistical fluctuations of $\hat{\mathbf{\Psi}}$. Denoting the **true** parameter covariance, i.e. the usual **inverse** Fisher matrix, as $\Sigma_{\mathbf{p}} = \mathbf{F}^{-1}$, we find the result [39]:

$$\langle \hat{\Sigma}_{\mathbf{p}} \rangle = \Sigma_{\mathbf{p}} \left[1 + \frac{N_b - N_p}{N_r} + \frac{(N_b - N_p)(N_b - N_p + 2)}{N_r^2} \right] + O\left(\frac{1}{N_r^3} \right). \quad (10)$$

Although we truncated the expansion to second order in $1/N_r$, an exact expression for $\langle \hat{\Sigma}_{\mathbf{p}} \rangle$ has been proposed by [40]

$$\langle \hat{\Sigma}_{\mathbf{p}} \rangle_{\rm empirical} = \Sigma_{\mathbf{p}} \left(\frac{N_r - 2}{N_r - N_b + N_p - 2} \right) \quad (11)$$

This empirical expression reduces to equation (10) when expanded at order $O(1/N_r^2)$ but, to our knowledge, no first principles proof of its correctness exists. Next, we restrict ourselves to the large N_r limit, and we further investigate the behavior of the $O(1/N_r)$ term. We consider three cases [41]:

1. If the true data covariance $\mathbf{C}$ is known, the estimator in eq. (8) is biased, and the dominant contribution of the bias comes from the second order fluctuations in $\hat{\mathbf{\Psi}}$. Once the expectation values over simulations are taken, the bias sums up to

$$\langle \hat{\Sigma}_{\mathbf{p}} \rangle = \Sigma_{\mathbf{p}} \left(1 + \frac{N_b - N_p}{N_r} \right). \quad (12)$$

   This is the result obtained by ref. [1].

2. Usually the true data covariance is unknown, and it is tempting to plug in its estimator $\hat{\mathbf{C}}$, measured from the same simulation set we use to compute $\hat{\mathbf{\Psi}}$. This approach has been used before in the literature (e.g. [17, 24]). If this is done without correcting for the bias in $\hat{\mathbf{\Psi}}$ (see Ref. [29] and eq. 20 below), the parameter variance will have a contribution from both the second and first-order fluctuations in $\hat{\mathbf{\Psi}}$, which now have a nonzero expectation value. In

Ω_m	w	σ_8	(N_s, N_r)	Number of κ ensembles
0.26	-1	0.8	(1 to 200,1024)	16
0.26	-1	0.8	(1,128000)	1
0.29	-1	0.8	(1,1024)	1
0.26	-0.8	0.8	(1,1024)	1
0.26	-1	0.6	(1,1024)	1

TABLE I. Summary of the shear ensembles used in this work. N_s and N_r refer to the number of independent N–body simulations, and the number of pseudo-independent realizations created from these simulations, respectively.

this case the bias sums up to

$$\langle\hat{\Sigma}_{\mathbf{p}}\rangle = \Sigma_{\mathbf{p}}\left(1 - \frac{N_b - N_p}{N_r}\right). \tag{13}$$

3. If we repeat the same exercise as above, but we correct for the bias in the precision matrix estimator, we are left with

$$\langle\hat{\Sigma}_{\mathbf{p}}\rangle = \Sigma_{\mathbf{p}}\left(1 + \frac{1 + N_p}{N_r}\right). \tag{14}$$

The error degradation in each parameter p, at leading order, scales as D/N_r, where $D = N_b - N_p$, $N_p - N_b$, and $1 + N_p$ for cases 1, 2, and 3, respectively. Note that in the last case, which is most relevant when fitting actual data, the estimated degradation turns out to be too optimistic: the parameter estimate $\hat{\mathbf{p}}$ has a variance whose noise grows linearly with N_b (eq. 10), whereas the degradation estimated via eq. (14) is constant with N_b. This can lead to underestimation of error bars, which can be mistakenly interpreted as a parameter bias. We test scaling relations of the form

$$\langle\hat{\sigma}_p^2\rangle = \sigma_{p,\infty}^2(N_s)\left(1 + \frac{D}{N_r}\right) \tag{15}$$

against our simulations, in the limits of both high and low N_r. We indicate the diagonal elements of $\hat{\Sigma}_{\mathbf{p}}$ as $\hat{\sigma}_p^2 = \mathrm{diag}(\hat{\Sigma}_{\mathbf{p}})$ and we indicate by $\sigma_{p,\infty}^2$ the expectation value of the variance of each parameter in the limit of an infinite number of realizations $N_r \to \infty$. We call D the *effective dimensionality* of the feature space (which, as seen before, can be negative in some pathological cases). We compute the expectation values of $\hat{\sigma}_p^2$ (eqs. 8 and 15) by averaging over 100 random resamplings of our shear ensembles. For the true feature covariance matrix $\langle(\hat{\mathbf{d}} - \mathbf{d}_0)(\hat{\mathbf{d}} - \mathbf{d}_0)^T\rangle = \mathbf{C}$ we use the estimated covariance from a grand ensemble built with the union of all the ensembles with different N_s.

The true parameter variance $\sigma_{p,\infty}^2(N_s)$ in principle can depend on the number of independent N–body simulations N_s, which appears in the randomization procedure described in § II A above. This is because if N_s is not large enough, the different shear realizations cannot be all independent, and hence the true variance $\sigma_{p,\infty}^2(N_s \to \infty)$ cannot be recovered for low N_s even if N_r is arbitrarily large. In the next section, we present our main findings.

III. RESULTS

In this section we present the main results of this work. We show the qualitative behavior of a variety of feature $\hat{\mathbf{d}}_r$ probability distribution functions (PDFs) in ensembles built with different N_s and N_r. In Figure 1, we show the PDF of the power spectrum at four selected multipoles, spanning the linear ($\ell = 115$) to the nonlinear ($\ell = 5283$) regime. In Figure 2, we shows the ensemble mean for these power spectra, as well as for peak counts of three different κ_0 heights (corresponding to $\approx 2 - 13\sigma$ peaks), as a function of N_s. In Figure 3, we show the variance of the power spectrum at each multipole, as a function of N_s, in units of the variance expected if the convergence κ was a Gaussian random field

$$\mathrm{Var}(P_\ell^{\kappa\kappa}) = \frac{(P_\ell^{\kappa\kappa})^2}{N_{\rm eff}(\ell).} \tag{16}$$

Here $N_{\rm eff}(l)$ is the number of independent modes used to estimate the power spectrum at ℓ.

In practice, we measure $P_\ell^{\kappa\kappa}$ on the Fourier transform of the pixelized simulated map $\kappa_r(\boldsymbol{\theta})$, using the FFT algorithm, and some care must be taken to count the number of modes $N_{\rm eff}(\ell)$ correctly. Each pixel (i_x, j_y) in Fourier space corresponds to a mode $(\ell_x, \ell_y) = 2\pi(i_x, i_y)/\theta_{\rm box}$, with $i_x = -n_{\rm ray}/2, ..., n_{\rm ray}/2$ and $i_y = 0, ..., n_{\rm ray}/2$. Here $n_{\rm ray} = 2048$ is the linear number of pixels on the ray–traced convergence maps. We count the number of pixels $N(\ell)$ that fall inside a multipole bin (ℓ_1, ℓ_2). Because the κ field is real, the modes $(\pm\ell_x, 0)$ are not independent. If we let $N(\ell, \ell_y = 0)$ be the number of non–independent modes, the effective number of independent modes for the variance is given by

$$N_{\rm eff}(\ell) = \frac{N^2(\ell)}{N(\ell) + N(\ell, \ell_y = 0)}. \tag{17}$$

This correction is important at low ℓ, where pixelization effects are non-negligible; $N_{\rm eff}(\ell \gg 2\pi/\theta_{\rm box}) \approx N(\ell)$.

In Figures 4 and 5, we show the dependence of the confidence range $\langle\hat{\sigma}_w^2\rangle$ on N_r, derived from the features used in this work (see Table II for a comprehensive list). Figure 4 shows the behavior in the limit of a large number $N_r \gg 500$ of realizations, and compares it with the scaling of the form in equation (15). Figure 5 shows the

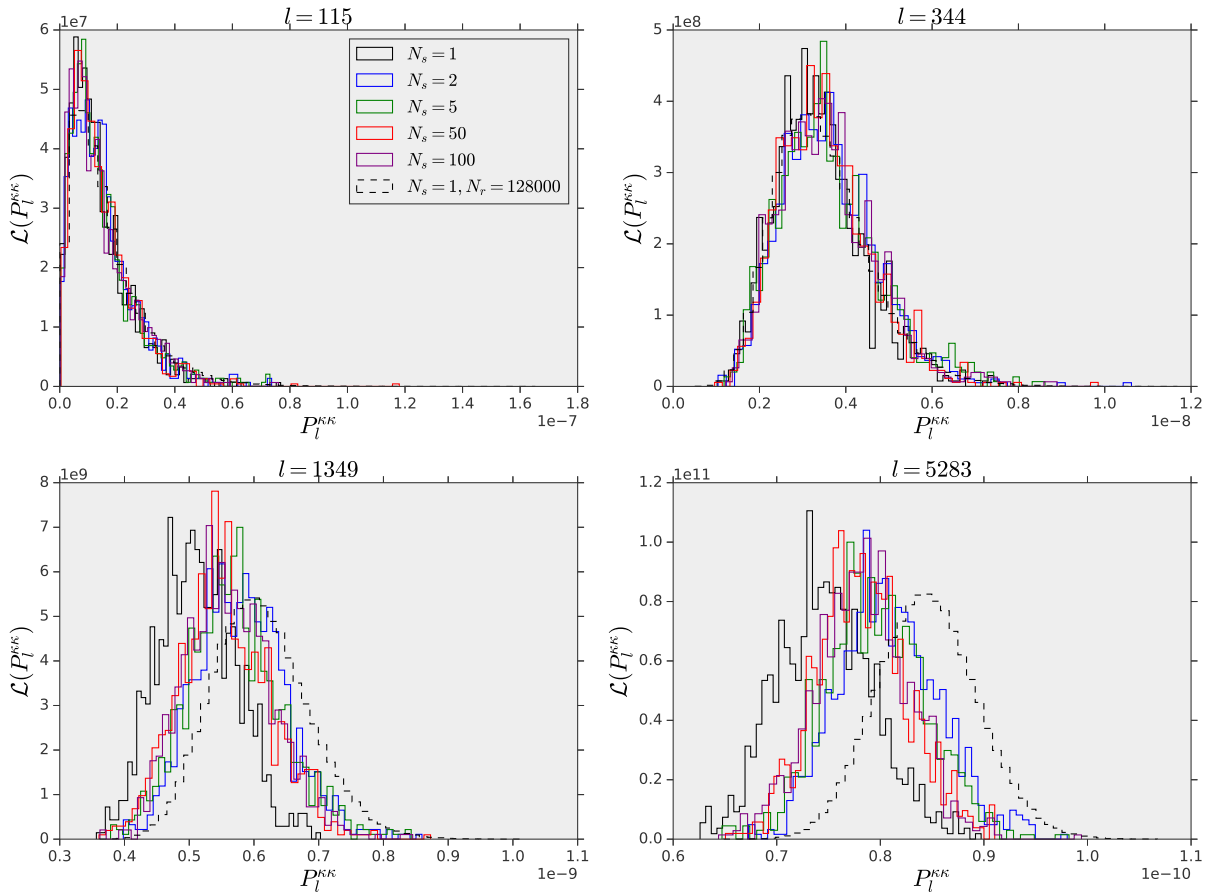

FIG. 1. PDF of the κ power spectrum $\mathcal{L}(P_l^{\kappa\kappa})$ at four selected multipoles $\ell = 115, 344, 1349, 5283$, for different shear ensembles constructed from on N_s =1 (black), 2 (blue), 5 (green), 50 (red), and 100 (purple) independent N–body simulations. Each curve is based on $N_r = 1024$ realizations. The dashed black curves correspond to ensembles generated with $N_s = 1$ and $N_r = 128000$. For $N_s \geq 2$, the distributions appear similar to the eye; this similarity is confirmed by the comparisons in Figures 2 and 3 below.

Feature	**Specifications**	N_b	**Symbol**	**Color**
Power Spectrum, log binning	$\ell \in [100, 800]$	8	×	black
Power Spectrum, log binning	$\ell \in [1000, 6000]$	7	■	black
Power Spectrum, log binning	$\ell \in [100, 6000]$	15	●	red
Power Spectrum, linear binning	$\ell \in [100, 2000]$	15	+	red
Power Spectrum, linear binning	$\ell \in [2500, 4500]$	15	×	red
Power Spectrum, linear binning	$\ell \in [100, 4500]$	30	●	green
Power Spectrum, linear binning	$\ell \in [100, 6000]$	39	●	blue
Low peaks	$\kappa_0 \in [-0.06, 0.09]$	15	+	red
Intermediate peaks	$\kappa_0 \in [0.1, 0.27]$	15	★	red
High peaks	$\kappa_0 \in [0.28, 0.45]$	15	◇	red
Low+Intermediate peaks	$\kappa_0 \in [-0.06, 0.27]$	30	×	green
Intermediate+High peaks	$\kappa_0 \in [0.1, 0.45]$	30	■	green
All peaks	$\kappa_0 \in [-0.06, 0.45]$	45	■	magenta

TABLE II. Catalog of feature types used in this work, along with the chosen number of bands N_b and the plot legends for Figure 4.

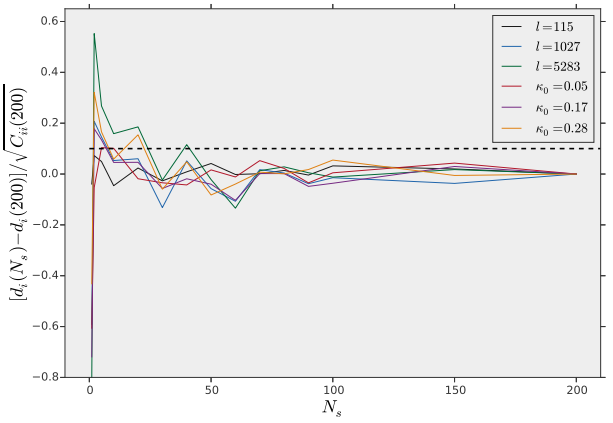

FIG. 2. The mean value $\bar{\mathbf{d}}$ of various features, measured from ensembles created from different numbers N_s of simulations. For each case, the difference compared to the mean in the $N_s = 200$ ensemble is shown, in units of the statistical error measured in the $N_s = 200$ ensemble. The colored curves refer to shear–shear power spectra measured at $\ell = 115$ (black), 1027 (cyan), and 5283 (green), and peak counts with heights $\kappa_0 = 0.05$ (red), 0.17 (purple), and 0.28 (orange). The κ bin width for the peak counts has been fixed to $\Delta\kappa = 0.011$. The dashed black line shows a level of 0.1σ accuracy for reference. For $N_s \geq 2$, the means are statistically indistinguishable (even at $\sim 0.1\sigma$) from those in the ensemble with $N_s = 200$.

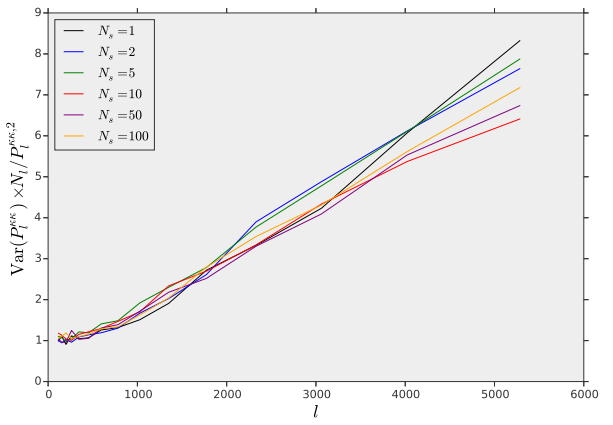

FIG. 3. Variance of the κ power spectrum as a function of the multipole ℓ, in units of the expected Gaussian variance from equation (16). The variance is measured from different shear ensembles based on N_s =1 (black), 2 (blue), 5 (green), 10 (red), 50 (purple), or 100 (orange) N–body simulations. Non–Gaussianities of the underlying structures increase the variance on small scales, but no clear trend with N_s can be identied on any scale.

large N_r trends of the w constraint. Figure 4 illustrates the behavior at relatively low N_r, and compares $\langle\hat{\sigma}_w^2\rangle$ measured directly from the simulations with the analytic expectations from equation (10). Finally, in Figure 6, we show how the w confidence limit changes with N_s.

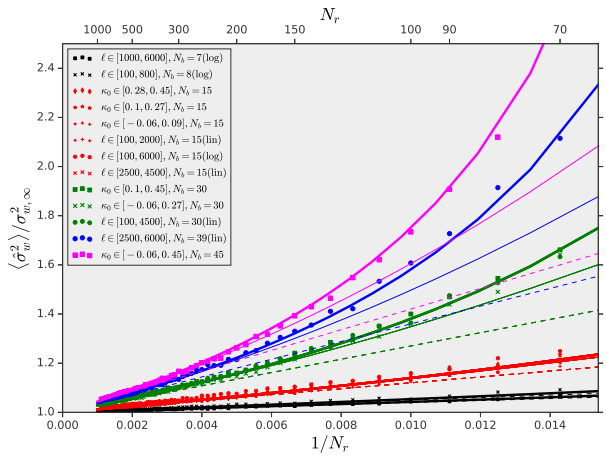

FIG. 4. Expectation value of the variance of w computed from equation (8), shown as a function of $1/N_r$. The different symbols and colors correspond to the features listed in Table II. The dashed and this solid curves show the analytic predictions from equation (10) at orders $O(1/N_r)$ and $O(1/N_r^2)$, respectively. **The thick solid curves show the empirical predictions from equation (11)**. The asymptotic variance $\sigma_{w,\infty}^2$ has been computed from a linear regression of $\langle\hat{\sigma}_w^2\rangle$ vs $1/N_r$ for $N_r > 500$. The figure clearly shows that terms beyond $O(1/N_r)$ need to be considered, unless $N_r \gg N_b$.

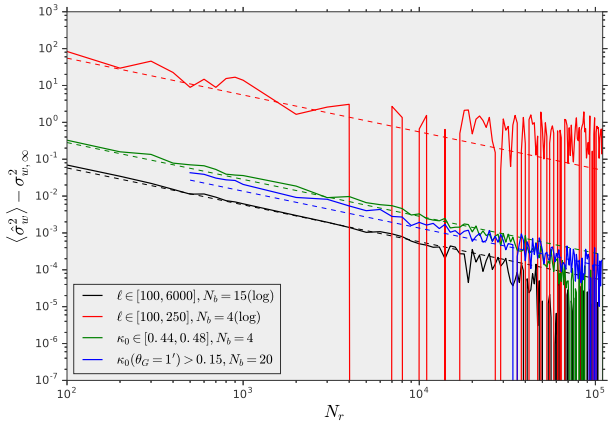

FIG. 5. Bias in the variance of w, $\langle\hat{\sigma}_w^2\rangle - \sigma_{w,\infty}^2$, as a function of the number of realizations N_r used to estimate the covariance (eq. 3). The figure shows both the trend measured in the simulations (solid lines) and their scaling expected from Eq. (15) with $D = N_b - N_p$ (dashed line). The asymptotic variance $\sigma_{w,\infty}^2$ has been estimated to be the value $\langle\hat{\sigma}_w^2\rangle(N_r = 10^5)$. Different features are considered: power spectra with logarithmically spaced $\ell \in [100, 6000]$ (black), $\ell \in [100, 250]$ (red), peak counts in the unsmoothed maps with height $\kappa_0 \in [0.44, 0.48]$ (green) and peak counts in the smoothed maps (with a Gaussian kernel of size $\theta_G = 1'$) with height $\kappa_0 > 0.15$ (blue). No deviations from the expected $1/N_r$ behavior are observed up to $N_r \approx$ few $\times\, 10^4$, except for the large-scale power spectrum, in which case the deviations occur much earlier ($N_r \approx 10^3$).

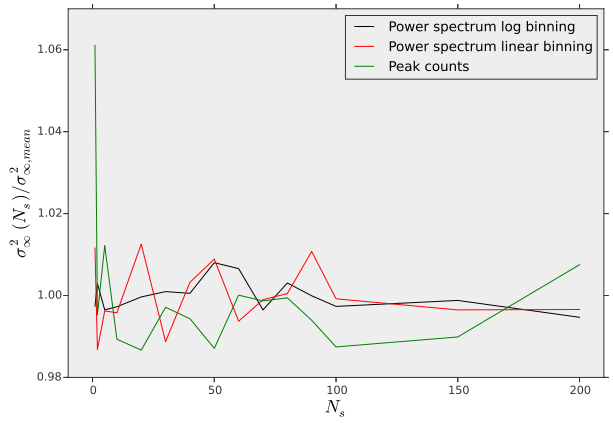

FIG. 6. The variance of w in the limit of $N_r \to \infty$ (measured from the intercept of the fit σ^2_w vs $1/N_r$), varying the number of simulations N_s used in the ensemble to estimate the covariance (eq. 3). We show the dependence of $\sigma^2_{w,\infty}(N_s)$ in units of the mean over the **union of** 16 ensembles with different N_s **and** $N_r = 1024$, for the power spectrum logarithmically binned (black, $N_b = 15, \ell \in [100, 6000]$), the power spectrum linearly binned (red,$N_b = 39, \ell \in [100, 6000]$) and the peak counts (green,$N_b = 45, \kappa_0 \in [-0.06, 0.45]$). No trend with N_s is seen for $N_s \geq 2$, and the differences are only of order 1%.

IV. DISCUSSION

In this section we discuss our main findings and their implications. Figure 1 shows that, although different choices of N_s do not affect the power spectrum PDF on large scales (top two panels), there are some qualitative differences on smaller scales (bottom two panels). On these smaller scales, shear ensembles built from $N_s = 1$ do not produce the same statistical behavior as ensembles built with larger N_s. In particular, looking at the black curves, we see that the $N_s = 1$ ensembles exibit large shifts with respect to the other PDFs to lower power, including the locations of the peaks of the PDFs. We attribute these offsets to large (random) statistical errors.

Interestingly, we need as few as $N_s = 2$ simulations to recover the right PDF for the small–scale power spectrum. Figure 2 shows that multiple independent N–body simulations $N_s \geq 2$ are indeed necessary for measuring the means of feature ensembles to an accuracy corresponding to 10% of the statistical error. The number of required simulations N_s depends on the feature type and ranges from a few ($N_s = 1$ or 2) for the power spectrum at low multipoles ($\ell \lesssim 500$) to $N_s \approx 30 - 50$ for the power spectrum at larger multipoles ($\ell \gtrsim 1000$) or peak counts above a high threshold ($\kappa_0 \approx 0.3$). On the other hand, relaxing the required accuracy to 50% of the statistical error, we find $N_s = 2$ to be always sufficient. **As pointed out by [42], the box size used for the N–body simulations can also play an important role in the accuracy of the power spectrum ensemble means.**

Figure 3 shows the variance of convergence power spectrum computed from different ensembles, in units of the Gaussian expectation. We find that, even with $N_s = 1$, we are able to recover the known result that non–Gaussian structures increase the variance significantly on small scales (see [34, 43] for reference). Our results are in fact in excellent quantitative agreement with [34], which used $N_s = 400$ independent N–body simulations. This result is highly encouraging, suggesting that individual N–body runs can be recycled repeatedly. However, it is not sufficient by itself to conclude that N_s does not impact the parameter inferences, since these depend on the cross band covariances.

Figure 4 investigates the parameter errors. This figure shows that error degradation estimates truncated at order $O(1/N_r)$ are too optimistic when the number of simulations N_r used to measure the covariance is only a factor of few larger than the dimension of the feature space N_b. In these cases effects coming the next–to–leading orders $O(1/N_r^2)$ become non–negligible on constraint degradation. In particular, we find that already for $N_b = 30$ and $N_r \sim 100$, the error degradation estimates to the next leading order, $O(1/N_r^2)$, remain too optimistic. Accurate analytic estimates in this regime would require at least terms of order $O(1/N_r^3)$, which come from higher–than–quartic $\hat{\mathbf{\Psi}}$ fluctations.

In Figure 5, we examine how the degradation in the w constraint depends on the number of simulations used to estimate the covariance, in the limit of large N_r. We find excellent agreement with the expected scaling (eq. 15) up to $N_r \sim$ few $\times 10^4$ when using the κ power spectrum in the multipole range $\ell \in [100, 6000]$. The same behavior is observed when considering the high-significance peak counts ($> 10\sigma$ for unsmoothed maps and $> 5\sigma$ for $1'$ smoothed maps). As the figure shows, around these values of N_r the $\langle\hat{\sigma}^2_w\rangle - \sigma^2_{w,\infty}$ curve becomes noisy and reaches negative values. This is a clear indication that the $1/N_r$ behavior is broken and a plateau in $\langle\hat{\sigma}^2_w\rangle$ is reached. The negative values in the plot are a consequence of the noise in the estimation of this plateau value (or equivalently in the estimated value of $\sigma^2_{w,\infty}$).

We conclude that a single N–body simulation is sufficient to construct an ensemble of up to a few$\times 10^4$ mutually independent convergence power spectra. For $N_r \gg 10^4$, the shear realizations can no longer be considered independent. We emphasize that the precise value of this N_r will depend on the size of the simulation box (which, in our case, is $(240\text{Mpc}/h)^3$, with 512^3 particles) and also on the range of multipoles ℓ used to constrain the parameters. Figure 5 shows that when we infer w only from large–scale modes, $\ell \lesssim 250$, the plateau is reached at least an order of magnitude earlier in the number of realizations. In other words, the number of independent power spectra we can generate decreases as we increase the spatial scales of interest. This is due to the fact that, because of the finite box size, the number of independent lens plane shifts (as described in § II A) decreases as the mode size approaches the size of the box. Similarly, one may expect that the independence in the

statistics of high-amplitude peaks, which are predominantly produced by single massive halos, may be compromised by these halos being present repeatedly, in many of the pseudo-independent realizations. However, Figure 5 shows that this is not the case: the peak count statistics are shown at κ thresholds corresponding to massive ($\approx 10^{15}$ M$_\odot$) halos, yet there is no evidence that the independence of the maps breaks down until N_r =few$\times 10^4$. Apparently, randomly projected structures, which vary from realization-to-realization, contribute significantly to the statistics of these high peaks.

Figure 6 shows how the "true" w constraint (in the limit $N_r \to \infty$; or equivalently the w constraint with the known N_r dependence factored out), depends on N_s. We find that, in the range $N_s \in [1, 200]$ the inferred w–variance $\sigma^2_{w,\infty}$ fluctuates stochastically only by 1%, and does not show any trend with N_s.

Finally, we found that when we estimate the data covariance $\mathbf{C}$ from the same simulation set used to measure $\hat{\mathbf{\Psi}}$, the effective dimensionality D decreases with increasing N_b in the case where the $\hat{\mathbf{\Psi}}$ bias is not corrected (eq. 13). This N_b-dependence disappears when the bias is corrected (eq. 14). This fact that should be taken into consideration when forecasting parameter errors purely from simulations, as the errors will otherwise be underestimated. A similar conclusion was reached by [44] (although their paper did not address the impact of using the same simulation set for $\mathbf{C}$ and $\hat{\mathbf{\Psi}}$).

V. CONCLUSIONS

In this work, we have examined the effect of forecasting cosmological constraints based on shear ensembles generated from a finite number of N–body simulations. Our main results can be summarized as follows:

- When the feature covariance matrix is measured from simulations, parameter constraints are degraded. This degradation is appreciably larger than the $O(1/N_r)$ computed by [1] when the number of realizations N_r is only a factor of few larger than the feature vector size N_b.
- We can recycle a single 240Mpc/h N–body simulation to produce an ensemble of $O(10^4)$ shear maps whose small-scale power spectra and high-signficiance peak counts are statistically independent. The mean feature measured from a shear ensemble, though, could be inaccurate if only one N–body simulation is used.
- As few as one or two independent N–body simulations are sufficient to forecast w error bars to 1% accuracy, provided that a sufficiently large number N_r of realizations are used to measure feature covariances. In particular, provided that biases in the inverse covariance are corrected, percent–level forecasts require $N_r \gtrsim 100(N_b - N_p)$ realizations.
- Depending on the feature type used to constrain cosmology, a larger number of N–body simulations might be needed to measure accurate ensemble means to an accuracy corresponding 10% of the statistical error. If this accuracy requirement is relaxed to 50% of the statistical error, we find that as low as $N_s = 2$ simulations are sufficient for the feature types we consider in this work.

Future extensions of this work should involve extending our analysis to a larger set of cosmological parameters, and to more general feature spaces, such as the ones that characterize non–Gaussian statistics (e.g. including higher moments of the κ field, Minkowski Functionals, and higher-order κ correlators). While our results are highly encouraging, and suggest that a single N–body simulation can be recycled repeatedly, to produce as many as 10^4 independent shear power spectra or peak count histograms. In order to scale our results to large future surveys, such as LSST, it will be necessary to determine if our findings hold when challenged by larger and higher-resolution N–body simulations [45].

ACKNOWLEDGMENTS

We thank Lam Hui for useful discussions. The simulations in this work were performed at the NSF XSEDE facility, supported by grant number ACI-1053575, and at the New York Center for Computational Sciences, a cooperative effort between Brookhaven National Laboratory and Stony Brook University, supported in part by the State of New York. This work was supported in part by the U.S. Department of Energy under Contract Nos. DE-AC02-98CH10886 and DE-SC0012704, and by the NSF Grant No. AST-1210877 (to Z.H.) and by the Research Opportunities and Approaches to Data Science (ROADS) program at the Institute for Data Sciences and Engineering at Columbia University (to Z.H.).

[1] S. Dodelson and M. D. Schneider, Phys. Rev. D **88**, 063537 (2013), arXiv:1304.2593 [astro-ph.CO].
[2] C. Heymans, L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, P. Simon, C. Bonnett, J. Coupon, L. Fu, J. Harnois Déraps, M. J. Hudson, M. Kilbinger, K. Kuijken, B. Rowe, T. Schrabback, E. Semboloni, E. van Uitert, S. Vafaei, and M. Velander, MNRAS **427**, 146 (2012), arXiv:1210.0032 [astro-ph.CO].
[3] L. Miller, C. Heymans, T. D. Kitching, L. van Waerbeke, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, B. T. P. Rowe, J. Coupon, J. P. Dietrich, L. Fu,

J. Harnois-Déraps, M. J. Hudson, M. Kilbinger, K. Kuijken, T. Schrabback, E. Semboloni, S. Vafaei, and M. Velander, MNRAS **429**, 2858 (2013), arXiv:1210.8201 [astro-ph.CO].

[4] A. M. Koekemoer, H. Aussel, D. Calzetti, P. Capak, M. Giavalisco, J.-P. Kneib, A. Leauthaud, O. Le Fèvre, H. J. McCracken, R. Massey, B. Mobasher, J. Rhodes, N. Scoville, and P. L. Shopbell, ApJS **172**, 196 (2007), astro-ph/0703095.

[5] The Dark Energy Survey Collaboration, ArXiv Astrophysics e-prints (2005), astro-ph/0510346.

[6] LSST Dark Energy Science Collaboration, ArXiv e-prints (2012), arXiv:1211.0310 [astro-ph.CO].

[7] J. Amiaux, R. Scaramella, Y. Mellier, B. Altieri, C. Burigana, A. Da Silva, P. Gomez, J. Hoar, R. Laureijs, E. Maiorano, D. Magalhães Oliveira, F. Renk, G. Saavedra Criado, I. Tereno, J. L. Auguères, J. Brinchmann, M. Cropper, L. Duvet, A. Ealet, P. Franzetti, B. Garilli, P. Gondoin, L. Guzzo, H. Hoekstra, R. Holmes, K. Jahnke, T. Kitching, M. Meneghetti, W. Percival, and S. Warren, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442 (2012) p. 84420Z, arXiv:1209.2228 [astro-ph.IM].

[8] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, B. Macintosh, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, D. Benford, M. Hudson, W.-S. Jeong, Y. Mellier, W. Traub, T. Yamada, P. Capak, J. Colbert, D. Masters, M. Penny, D. Savransky, D. Stern, N. Zimmerman, R. Barry, L. Bartusek, K. Carpenter, E. Cheng, D. Content, F. Dekens, R. Demers, K. Grady, C. Jackson, G. Kuan, J. Kruk, M. Melton, B. Nemati, B. Parvin, I. Poberezhskiy, C. Peddie, J. Ruffa, J. K. Wallace, A. Whipple, E. Wollack, and F. Zhao, ArXiv e-prints (2015), arXiv:1503.03757 [astro-ph.IM].

[9] X. Yang, J. M. Kratochvil, K. Huffenberger, Z. Haiman, and M. May, Phys. Rev. D **87**, 023511 (2013), arXiv:1210.0608.

[10] E. Semboloni, H. Hoekstra, and J. Schaye, MNRAS **434**, 148 (2013), arXiv:1210.7303.

[11] M. White, Astroparticle Physics **22**, 211 (2004), astro-ph/0405593.

[12] H. Zhan and L. Knox, ApJ **616**, L75 (2004), astro-ph/0409198.

[13] A. R. Zentner, D. H. Rudd, and W. Hu, Phys. Rev. D **77**, 043507 (2008).

[14] D. H. Rudd, A. R. Zentner, and A. V. Kravtsov, ApJ **672**, 19 (2008), astro-ph/0703741.

[15] X. Yang, J. M. Kratochvil, S. Wang, E. A. Lim, Z. Haiman, and M. May, Phys. Rev. D **84**, 043529 (2011), arXiv:1109.6333.

[16] J. M. Kratochvil, E. A. Lim, S. Wang, Z. Haiman, M. May, and K. Huffenberger, Phys. Rev. D **85**, 103513 (2012), arXiv:1109.6334 [astro-ph.CO].

[17] A. Petri, Z. Haiman, L. Hui, M. May, and J. M. Kratochvil, Phys. Rev. D **88**, 123002 (2013), arXiv:1309.4460.

[18] L. Marian, R. E. Smith, and G. M. Bernstein, ApJ **698**, L33 (2009), arXiv:0811.1991.

[19] M. Takada and B. Jain, MNRAS **337**, 875 (2002), astro-ph/0205055.

[20] M. Takada and B. Jain, MNRAS **344**, 857 (2003), astro-ph/0304034.

[21] M. Takada and B. Jain, MNRAS **348**, 897 (2004), astro-ph/0310125.

[22] J. Bergé, A. Amara, and A. Réfrégier, ApJ **712**, 992 (2010), arXiv:0909.0529.

[23] J. P. Dietrich and J. Hartlap, MNRAS **402**, 1049 (2010), arXiv:0906.3512.

[24] M. Shirasaki, N. Yoshida, and T. Hamana, ApJ **774**, 111 (2013), arXiv:1304.2164 [astro-ph.CO].

[25] D. Bard, J. M. Kratochvil, C. Chang, M. May, S. M. Kahn, Y. AlSayyad, Z. Ahmad, J. Bankert, A. Connolly, R. R. Gibson, K. Gilmore, E. Grace, Z. Haiman, M. Hannel, K. M. Huffenberger, J. G. Jernigan, L. Jones, S. Krughoff, S. Lorenz, S. Marshall, A. Meert, S. Nagarajan, E. Peng, J. Peterson, A. P. Rasmussen, M. Shmakova, N. Sylvestre, N. Todd, and M. Young, ApJ **774**, 49 (2013), arXiv:1301.0830.

[26] C. Chang, S. M. Kahn, J. G. Jernigan, J. R. Peterson, Y. AlSayyad, Z. Ahmad, J. Bankert, D. Bard, A. Connolly, R. R. Gibson, K. Gilmore, E. Grace, M. Hannel, M. A. Hodge, M. J. Jee, L. Jones, S. Krughoff, S. Lorenz, P. J. Marshall, S. Marshall, A. Meert, S. Nagarajan, E. Peng, A. P. Rasmussen, M. Shmakova, N. Sylvestre, N. Todd, and M. Young, MNRAS **428**, 2695 (2013), arXiv:1206.1378.

[27] D. Huterer, M. Takada, G. Bernstein, and B. Jain, MNRAS **366**, 101 (2006), astro-ph/0506030.

[28] Sample variance is a broad term that has been used in the literature to describe a range of phenomena. Throughout this paper, we use it to refer to the fluctuations in an ensemble of simulations, which represent random realizations of the same initial conditions.

[29] A. Taylor, B. Joachimi, and T. Kitching, MNRAS **432**, 1928 (2013), arXiv:1212.4359.

[30] V. Springel, MNRAS **364**, 1105 (2005), astro-ph/0505010.

[31] S. Hilbert, J. Hartlap, S. D. M. White, and P. Schneider, A&A **499**, 31 (2009), arXiv:0809.5035.

[32] B. Jain, U. Seljak, and S. White, ApJ **530**, 547 (2000), astro-ph/9901191.

[33] A. Petri, "The lenstools computing package," (2015).

[34] M. Sato, T. Hamana, R. Takahashi, M. Takada, N. Yoshida, T. Matsubara, and N. Sugiyama, ApJ **701**, 945 (2009), arXiv:0906.2237 [astro-ph.CO].

[35] K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence, and C. Wagner, ApJ **705**, 156 (2009), arXiv:0902.0429 [astro-ph.CO].

[36] M. Kilbinger, K. Benabed, J. Guy, P. Astier, I. Tereno, L. Fu, D. Wraith, J. Coupon, Y. Mellier, C. Balland, F. R. Bouchet, T. Hamana, D. Hardin, H. J. McCracken, R. Pain, N. Regnault, M. Schultheis, and H. Yahagi, A&A **497**, 677 (2009), arXiv:0810.5129.

[37] M. D. Schneider, L. Knox, S. Habib, K. Heitmann, D. Higdon, and C. Nakhleh, Phys. Rev. D **78**, 063529 (2008).

[38] S. Matsumoto, ArXiv e-prints (2010), arXiv:1004.4717 [math.ST].

[39] The details of the calculation are given in Appendix A.

[40] A. Taylor and B. Joachimi, MNRAS **442**, 2728 (2014), arXiv:1402.6983.

[41] The details of these calculations are given in Appendix B.
[42] L. Casarini, O. F. Piattella, S. A. Bonometto, and M. Mezzetti, ApJ **812**, 16 (2015), arXiv:1406.5374.
[43] M. Takada and D. N. Spergel, MNRAS **441**, 2456 (2014), arXiv:1307.4399.
[44] J. Hartlap, P. Simon, and P. Schneider, A&A **464**, 399 (2007), arXiv:astro-ph/0608064.
[45] K. Heitmann, N. Frontiere, C. Sewell, S. Habib, A. Pope, H. Finkel, S. Rizzi, J. Insley, and S. Bhattacharya, ApJS **219**, 34 (2015), arXiv:1411.3396.

APPENDIX A: CUBIC AND QUARTIC COVARIANCE FLUCTUATIONS

The goal of this appendix is to give a derivation of eq. (10). When the simulated feature vector $\hat{\mathbf{d}}_r$ is drawn from a Gaussian distribution, the covariance estimator $\hat{\mathbf{C}}$ follows the Wishart distribution, and its inverse $\hat{\mathbf{\Psi}}$ follows the inverse Wishart distribution (see Ref. [29] for analytical expressions for these probability distributions). Computing expectation values of eq. (8) over the inverse Wishart distribution is not possible analytically, and a perturbative expansion is necessary. Writing $\hat{\mathbf{\Psi}} = \mathbf{\Psi} + \delta\hat{\mathbf{\Psi}}$, we can expand eq. (8) in powers of $\delta\hat{\mathbf{\Psi}}$. The expectation value of each term in this expansion can be calculated in terms of moments of the inverse Wishart distribution. Ref. [38] provides a general framework to compute these moments, and give exact expressions for moments up to quartic order. First, let us expand the inverse of the Fisher matrix estimator (eq. 9) in powers of $\delta\hat{\mathbf{\Psi}}$. The n–th order of this expansion will be

$$\delta\hat{\mathbf{F}}^{-1}_{(n)} = (-1)^n(\mathbf{F}^{-1}\delta\hat{\mathbf{F}})^n\mathbf{F}^{-1} \tag{18}$$

with

$$\delta\hat{\mathbf{F}} = \mathbf{d}_0'^T\delta\hat{\mathbf{\Psi}}\mathbf{d}_0' \tag{19}$$

Using eq. (18), we can expand eq. (8) to an arbitrary order in $\delta\hat{\mathbf{\Psi}}$, take the expectation values of the fluctuations over the inverse Wishart distribution, and finally arrive at eq. (10). We use the notation $\nu \equiv N_r - 1$ and $\gamma \equiv (\nu - N_b - 1)/2$, and we indicate with capital letters pairs of matrix indices, for example $I = (i_1, i_2)$, where $i_a = 1..N_b$. The main results we utilize from ref. [38] regarding the first four moments are (up to order $O(1/\nu^2)$)

$$\langle\hat{\Psi}_I\rangle = \frac{\nu}{2\gamma}\Psi_I \tag{20}$$

$$\langle\delta\hat{\Psi}_I\delta\hat{\Psi}_J\rangle = \frac{\nu^2\Psi_I\Psi_J + \nu^2\gamma\Psi_{\{I}\Psi_{J\}}}{4\gamma^2(\gamma-1)(2\gamma+1)} \tag{21}$$

$$\langle\delta\hat{\Psi}_I\delta\hat{\Psi}_J\delta\hat{\Psi}_K\rangle = \frac{\nu^3\Psi_{\{I}\Psi_J\Psi_{K\}}}{8\gamma(\gamma-1)(\gamma-2)(\gamma+1)(2\gamma+1)} \tag{22}$$

$$\langle\delta\hat{\Psi}_I\delta\hat{\Psi}_J\delta\hat{\Psi}_K\delta\hat{\Psi}_L\rangle = \frac{\nu^4(2\gamma^2-5\gamma+9)\Psi_{\{I}\Psi_{J\}}\Psi_{\{K}\Psi_{L\}}}{16\gamma(\gamma-1)(\gamma-2)(\gamma-3)(2\gamma-1)(\gamma+1)(2\gamma+1)(2\gamma+3)}. \tag{23}$$

Here the curly bracket notation is a shorthand for a symmetrization over pair of indices: for example

$$\Psi_{\{I}\Psi_{J\}} = \Psi_{i_1j_1}\Psi_{i_2j_2} + \Psi_{i_1j_2}\Psi_{i_2j_1} \tag{24}$$

Eq. (20) expresses the bias in the $\hat{\mathbf{\Psi}}$ estimator that already appears in the literature [44]. If we want to use the bias-corrected $\hat{\mathbf{\Psi}}$ estimator (required for the perturbative expansion of eq. 8), we need to apply an additional factor of $(2\gamma/\nu)^n$ to eqs. (20–23), where n is the order of the moment up to which we are applying the correction. If we limit ourselves to computing the expectation value of eq. (8) up to order $O(1/\nu^2)$, we do not need to worry about this correction for eqs. (22–23), as the dominant term here is already $O(1/\nu^2)$. The next step is expanding eq. (8) in powers of $\delta\hat{\mathbf{\Psi}}$ up to fourth order: this is easily done:

$$\hat{\Sigma}_{\mathbf{p}} = \left(\mathbf{F}^{-1} + \sum_{n=1}^{4}\delta\hat{\mathbf{F}}^{-1}_{(n)}\right)\mathbf{d}_0'^T(\mathbf{\Psi}+\delta\hat{\mathbf{\Psi}})\mathbf{C}(\mathbf{\Psi}+\delta\hat{\mathbf{\Psi}})\mathbf{d}_0'\left(\mathbf{F}^{-1} + \sum_{n=1}^{4}\delta\hat{\mathbf{F}}^{-1}_{(n)}\right). \tag{25}$$

Carrying out the calculations is simpler than it looks: because of the structure of eq. (25), each term in the expansion is proportional to $\Sigma_{\mathbf{p}} f_a(N_b, N_p)/N_r^a$, where

$f_a(N_b, N_p)$ is a polynomial in N_b and N_p. Terms proportional to N_b arise from index contractions of type $\mathrm{tr}(\mathbf{\Psi C}) = N_b$, which come from symmetrization terms of type $\Psi_{\{I}\Psi_{J\}}$. Symmetrization factors of type $\Psi_{\{I}\Psi_J\Psi_{K\}}$ or $\Psi_{\{I}\Psi_{J\}}\Psi_{\{K}\Psi_{L\}}$, on the other hand, give rise to contractions of type $\mathrm{tr}(\mathbf{\Psi C})\mathrm{tr}(\mathbf{F F^{-1}}) = N_b N_p$. Moreover, we know that, at every order $O(1/N_r^a)$, $f_a(N_b, N_p)$ has to be proportional to $N_b - N_p$, because it must vanish when $N_b = N_p$. The reason for this is that if the feature derivative matrix $\mathbf{d}_0'$ is square and invertible (which it should be in absence of degeneracies), then eq. (8) reduces to

$$\hat{\Sigma}_{\mathbf{p}} = (\mathbf{d}_0')^{-1}\mathbf{C}(\mathbf{d}_0'^T)^{-1}. \tag{26}$$

Every trace of the noise is gone, hence powers of $1/N_r^a$ must not appear at any order if $N_b = N_p$. Armed with the knowledge of the above considerations, we can compute the expectation value of eq. (25) at second, third and fourth order in $\delta\hat{\mathbf{\Psi}}$, keeping the terms that are at most $O(1/\nu^2) = O(1/N_r^2)$. When the combinatorial factors that arise from the expansion of eq. (25) are properly computed and the expectation values over the inverse Wishart distribution are taken according to eqs. (20–23), the results take the form

$$\begin{cases} (\delta\hat{\mathbf{\Psi}})^2 \to \Sigma_{\mathbf{p}} \dfrac{\gamma(N_b - N_p)}{(\gamma-1)(2\gamma+1)} = \Sigma_{\mathbf{p}} \left[\dfrac{N_b - N_p}{N_r} + \dfrac{(N_b - N_p)(N_b+3)}{N_r^2} \right] \\ (\delta\hat{\mathbf{\Psi}})^3 \to -4\Sigma_{\mathbf{p}} \dfrac{(N_b - N_p)(1+N_p)}{N_r^2} \\ (\delta\hat{\mathbf{\Psi}})^4 \to 3\Sigma_{\mathbf{p}} \dfrac{(N_b - N_p)(1+N_p)}{N_r^2}. \end{cases} \tag{27}$$

When the results from eq. (27) are summed, eq. (10) immediately follows.

APPENDIX B: NEGATIVE EFFECTIVE DIMENSIONALITY

The goal of this appendix is to give a justification for why the effective dimensionality D that appears in eq. (15) can be negative in some cases. When we use the same simulation set to estimate $\mathbf{C}, \hat{\mathbf{\Psi}}$, eq. (8) reduces to the inverse Fisher estimator $\hat{\Sigma}_{\mathbf{p}} = \hat{\mathbf{F}}^{-1}$. At second order in the Ψ fluctuations this becomes

$$\hat{\Sigma}_{\mathbf{p}} = \mathbf{F}^{-1} + \mathbf{F}^{-1}\left(-\delta\hat{\mathbf{F}} + \delta\hat{\mathbf{F}}\mathbf{F}^{-1}\delta\hat{\mathbf{F}}\right)\mathbf{F}^{-1} \tag{28}$$

If the biased estimator for $\hat{\mathbf{\Psi}}$ is used, we can use eqs. (20–21) at order $O(1/\nu)$ to compute

$$\langle\hat{\Sigma}_{\mathbf{p}}\rangle = \Sigma_{\mathbf{p}}\left(1 - \frac{N_b+1}{N_r} + \frac{1+N_p}{N_r}\right) = \Sigma_{\mathbf{p}}\left(1 + \frac{N_p - N_b}{N_r}\right). \tag{29}$$

We immediately see that the coefficient of $1/N_r$ is negative, because $N_b > N_p$. This is the result shown in eq. (13). If the bias correction for $\hat{\mathbf{\Psi}}$ is applied, the first order terms $\delta\hat{\mathbf{F}}$ average to 0, and we are left with only the last term in the sum eq. (29), which immediately yields eq. (14).